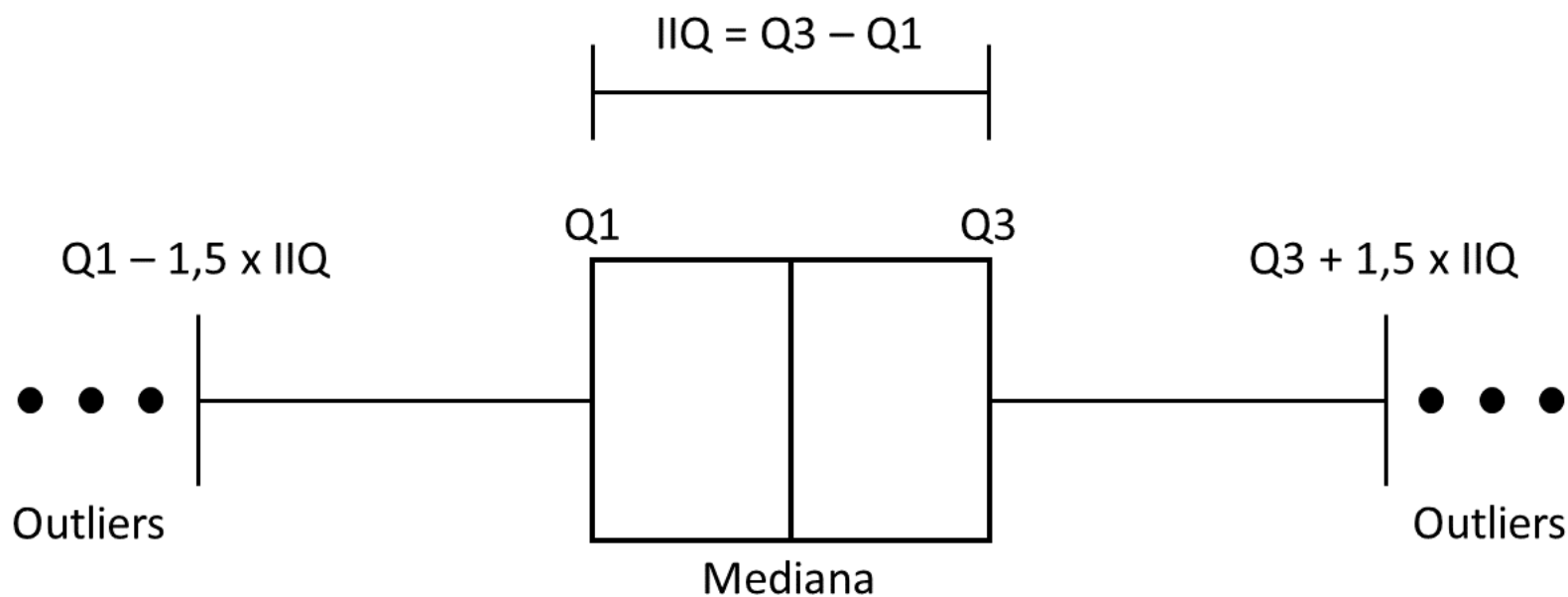


Transcrição

Veremos o **boxplot** neste passo, o qual é construído a partir de Medidas Separatrizes.

Na parte "4.2 Box-plot" de nosso notebook, encontraremos a representação gráfica que fornece muitas informações sobre nossos dados, como posição, dispersão, assimetria, discrepâncias e etc.



Box-plot

O "Q1" representa o primeiro quartil de 25%, teremos a mediana ao meio e o terceiro quartil representado por "Q3". A diferença entre os dois é o **índice interquartil** "I!Q".

Os extremos da representação gráfica são construídos como " $Q1 - 1,5(I\!I\!Q)$ " e " $Q3 + 1,5(I\!I\!Q)$ " respectivamente.

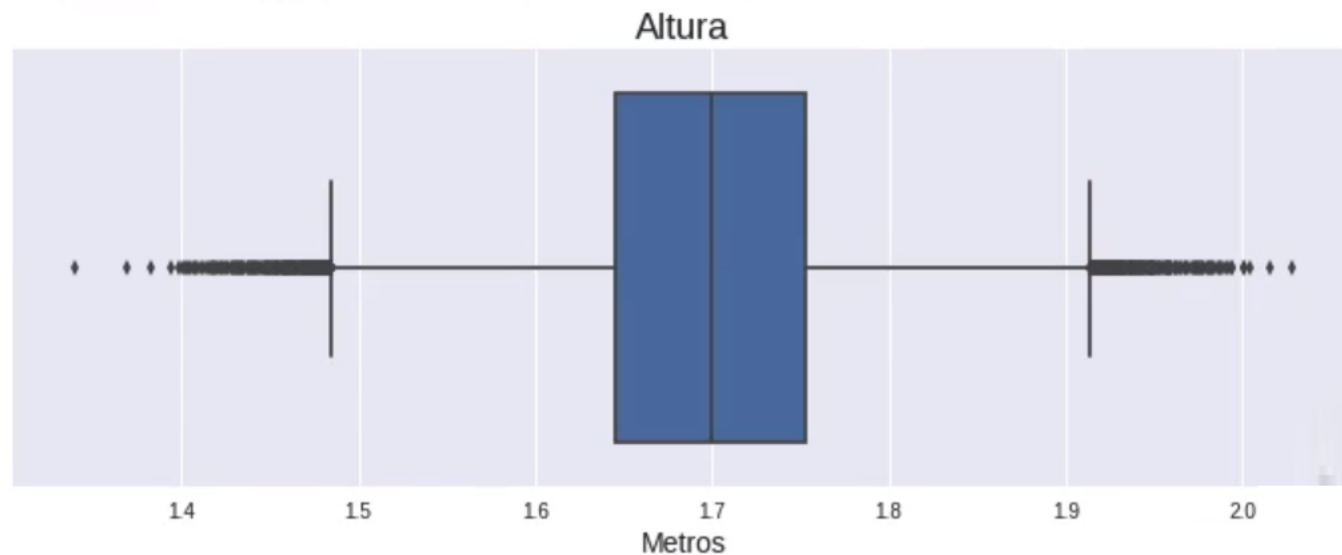
O "IIQ" é uma informação de **dispersão de dados**. Abaixo do primeiro limite e após o segundo, teremos candidatos a **dados discrepantes**, o que não significa que são *outliers* de imediato; como pesquisadores, teremos informações com calma e decidir se são de fato ou não.

Na primeira célula, criaremos o `.boxplot()` com `sns`. Passaremos inicialmente três parâmetros: `x` igual a `'Altura'` que é uma variável mais simétrica, o `data` será `dados` e `orient` será igual a `'h'` para que seja exibida na posição horizontal.

Em seguida, configuraremos as polegadas, os títulos, as labels, e os tamanhos de fonte.

```
ax = sns.boxplot( x = 'Altura', data = dados, orient = 'h')
ax.figure.set_size_inches(12, 4)
ax.set_title('Altura', fontsize=18)
ax.set_xlabel('Metros', fontsize=14)
ax
```

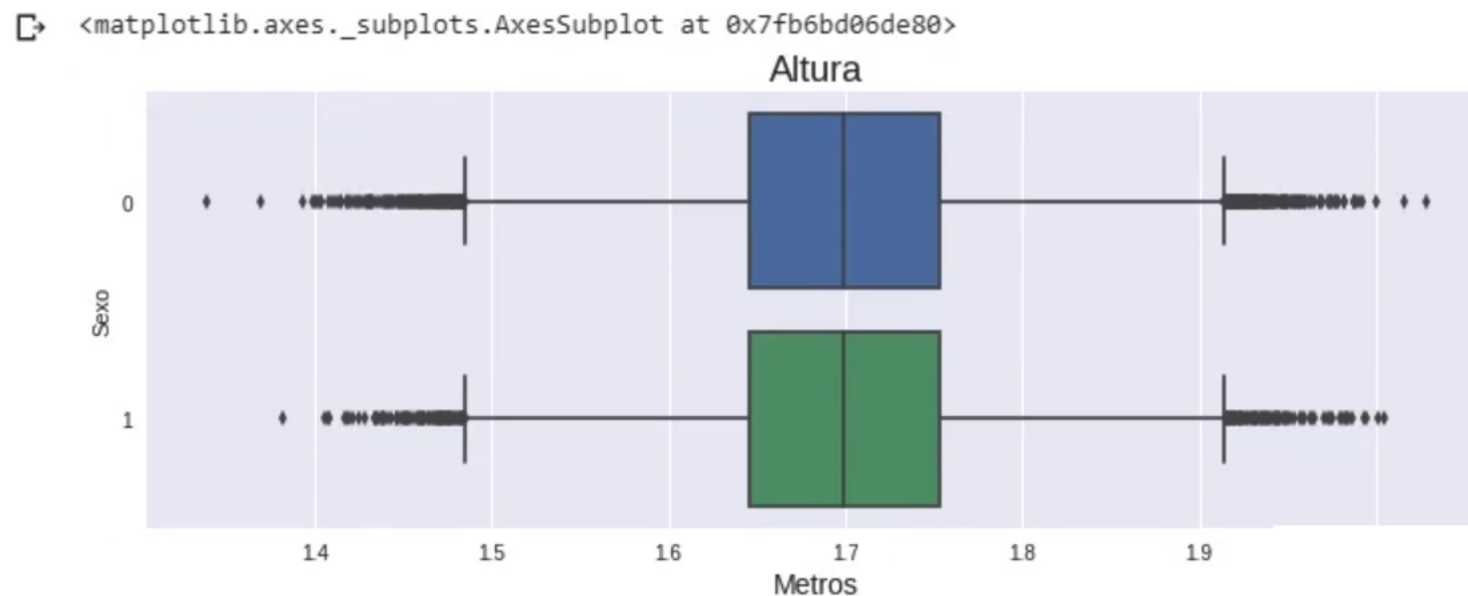
📄 <matplotlib.axes._subplots.AxesSubplot at 0x7fb6bd0f4e80>



Com este resultado, veremos que se trata de Simetria.

Esta construção nos permite fazer a comparação com outra variável, como `'Sexo'` no eixo `y`.

```
ax = sns.boxplot( x = 'Altura', y = 'Sexo', data = dados, orient = 'h')
ax.figure.set_size_inches(12, 4)
ax.set_title('Altura', fontsize=18)
ax.set_xlabel('Metros', fontsize=14)
ax
```



Não veremos quase nenhuma diferença entre as variáveis, pois `Altura` foi construída simetricamente para este curso, então os gráficos não revelam fatos da realidade da pesquisa.

Faremos a mesma coisa para a `Renda`, então poderemos apenas copiar o código da célula anterior e alterar algumas informações.

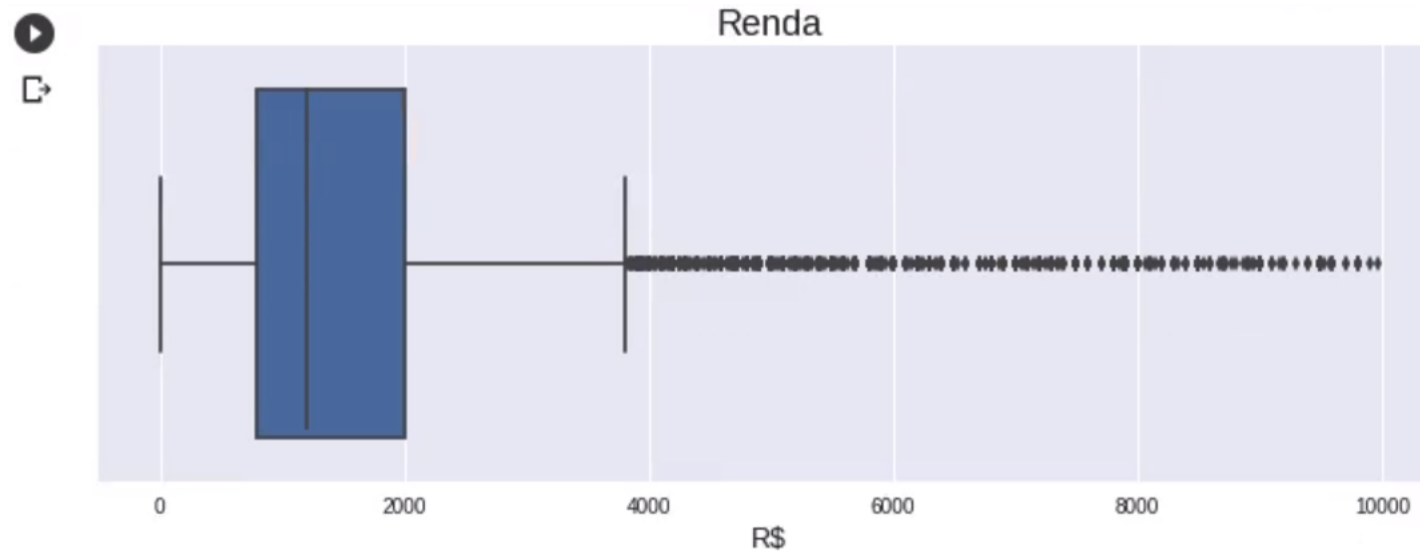
```
ax = sns.boxplot( x = 'Renda', data = dados, orient = 'h')
ax.figure.set_size_inches(12, 4)
ax.set_title('Renda', fontsize=18)
ax.set_xlabel('R$', fontsize=14)
ax
```



Neste gráfico, poderemos notar a enorme desigualdade social no Brasil, pois mal conseguimos visualizar a assimetria próxima de zero no `boxplot` .

Logo, selecionaremos apenas um grupo de pessoas que ganham menos de R\$10.000,00 dentro da variável, e o faremos por meio de `query()` recebendo `'Renda < 10000'` .

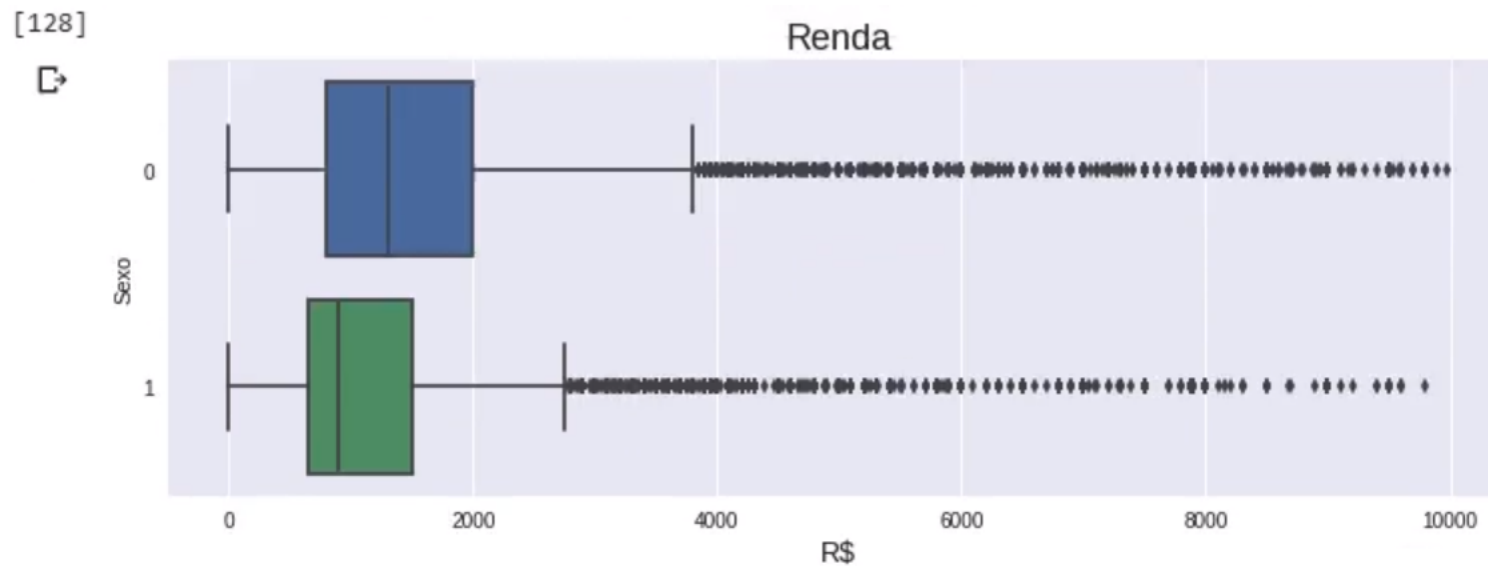
```
ax = sns.boxplot( x = 'Renda', data = dados.query('Renda < 10000'), orient = 'h')
ax.figure.set_size_inches(12, 4)
ax.set_title('Renda', fontsize=18)
ax.set_xlabel('R$', fontsize=14)
ax
```



Com isso, teremos uma visualização mais clara, ainda que bastante assimétrica mesmo com a seleção.

Faremos o mesmo código para visualizarmos a `Renda` com esta mesma `query()`. Porém, como eixo `y`, cruzaremos com a variável `'Sexo'`.

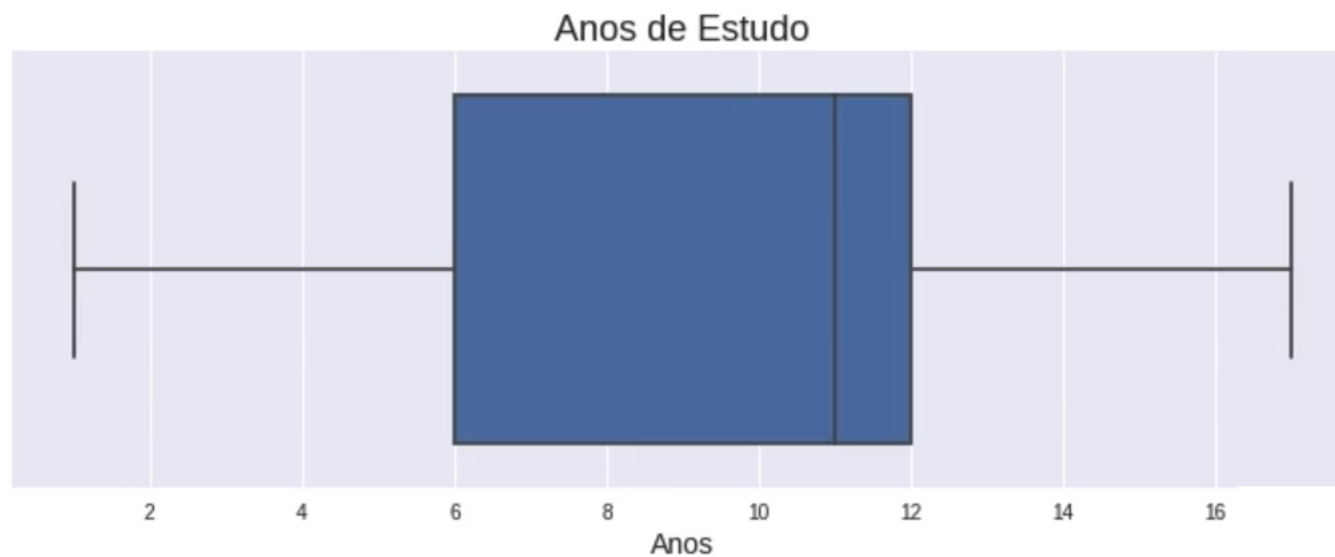
```
ax = sns.boxplot( x = 'Renda', y = 'Sexo', data = dados.query('Renda < 10000'), orient = 'h')
ax.figure.set_size_inches(12, 4)
ax.set_title('Renda', fontsize=18)
ax.set_xlabel('R$', fontsize=14)
ax
```



Como resultado da breve análise deste gráfico, veremos a grande diferença de renda entre homens e mulheres.

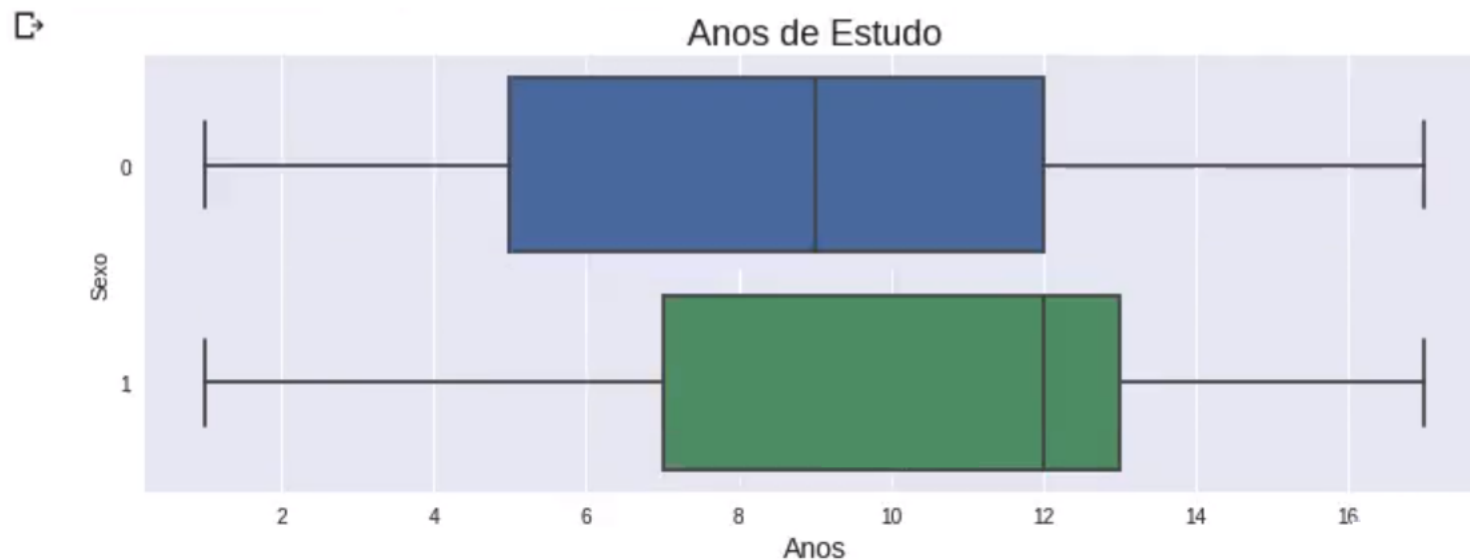
Na célula seguinte, faremos o `boxplot()` da variável 'Anos de Estudo'. Para fins de estudo, assumiremos que se tratam de anos e não uma classe.

```
ax = sns.boxplot( x = 'Anos de Estudo', data = dados, orient = 'h')
ax.figure.set_size_inches(12, 4)
ax.set_title('Anos de Estudo', fontsize=18)
ax.set_xlabel('Anos', fontsize=14)
ax
```



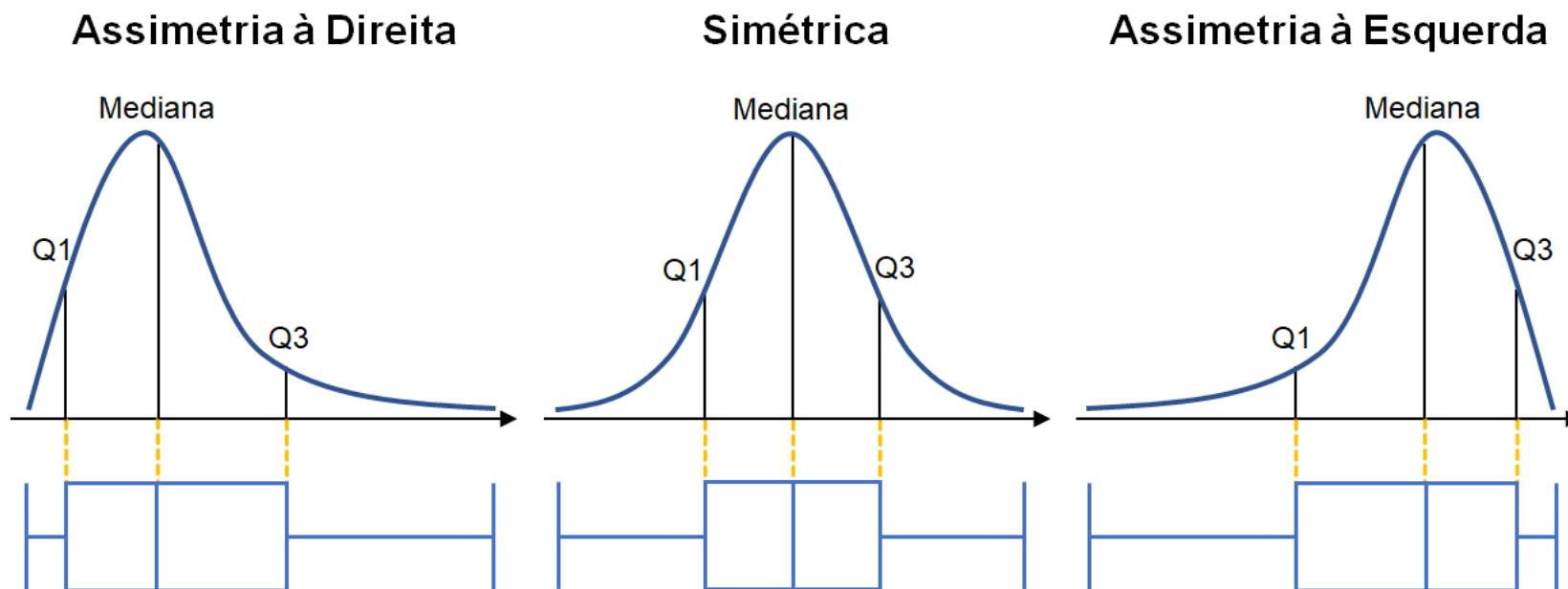
Observando o gráfico, veremos uma Assimetria à Esquerda. Em seguida, cruzaremos com a variável `Sexo` para constataremos se há alguma diferença de estudos entre homens e mulheres de fato também.

```
ax = sns.boxplot( x = 'Anos de Estudo', y = 'Sexo', data = dados, orient = 'h')
ax.figure.set_size_inches(12, 4)
ax.set_title('Anos de Estudo', fontsize=18)
ax.set_xlabel('Anos', fontsize=14)
ax
```



Neste resultado, poderemos comparar com o gráfico de renda dividido entre os sexos. Facilmente concluiremos que, apesar de as mulheres possuírem mais tempo de estudo, sua renda é inferior às dos homens.

Também poderemos notar a relação entre os gráficos `boxplot` e os de Assimetria à Direita, Simetria e Assimetria à Esquerda, pois possuem os mesmos comportamentos.



Logo, comparando os gráficos de simetria com os resultados deste passo, constataremos as semelhanças de características das variáveis e conseguiremos tirar algumas conclusões.

A seguir, falaremos sobre as **Medidas de Dispersão**.